Data Mining Exam

March 20, 2021

Kasper Rosenkrands

Aalborg University
Denmark

AALBORG UNIVERSITY
DENMARK

Agenda

Clustering
- Introduction
- K-Means
- Implementation of K-Means
- An example of the K-Means algorithm
- Hierarchical Clustering

Shrinkage

Classification

Trees

Support Vector Machines

Neural Networks

Clustering
Introduction

- **Clustering** is a way to categorize data to impose structure.
- A use case is recommender systems (Amazon, Spotify, Netflix), where a user is recommended items that bought/listened to/watched by other users with similar interests.

Clustering
K-Means

Given $D = (x_1, \ldots, x_n)$ where $x_i \in \mathbb{R}^p$, $K \in \mathbb{N}$ and let $C_1, \ldots, C_K$ denote different groups of the x_i's.

The K-Means algorithm tries to solve

$$\min_{C_1,\ldots,C_K} \left\{ \sum_{k=1}^{K} W(C_k) \right\}, \tag{1}$$

where $W(C_k)$ denotes the **within cluster variation**, in other words the dissimilarity of the group.

The most common dissimilarity measure is the is the squared Euclidean distance

$$W(C_k) := \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{i,j} - x_{i',j})^2. \tag{2}$$

Clustering
K-Means

You have reaced slide 18 lecture 2, keep going from there!

Clustering
Implementation of K-Means

Clustering
An example of the K-Means algorithm

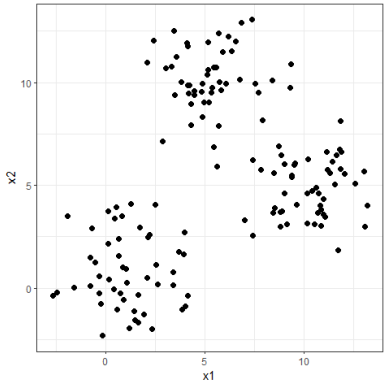

Figure: Iteration 01

Clustering
An example of the K-Means algorithm

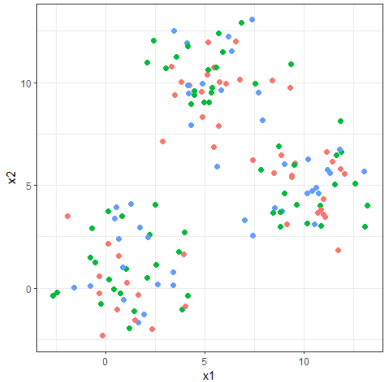

Figure: Iteration 02

Clustering
An example of the K-Means algorithm

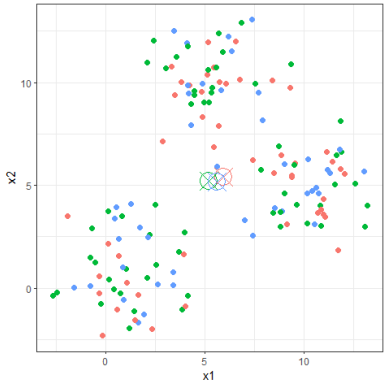

Figure: Iteration 03

Clustering
An example of the K-Means algorithm

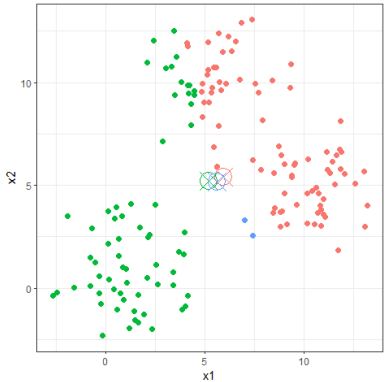

Figure: Iteration 04

Clustering
An example of the K-Means algorithm

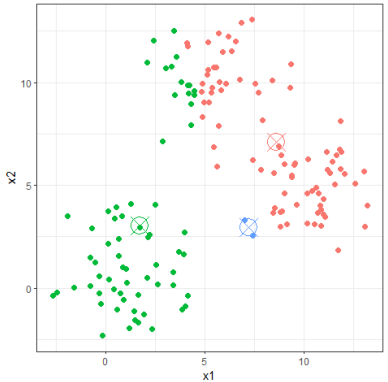

Figure: Iteration 05

Clustering
An example of the K-Means algorithm

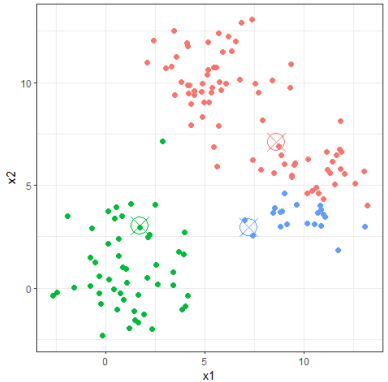

Figure: Iteration 06

Clustering
An example of the K-Means algorithm

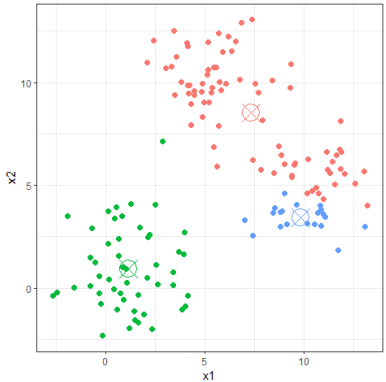

Figure: Iteration 07

Clustering
An example of the K-Means algorithm

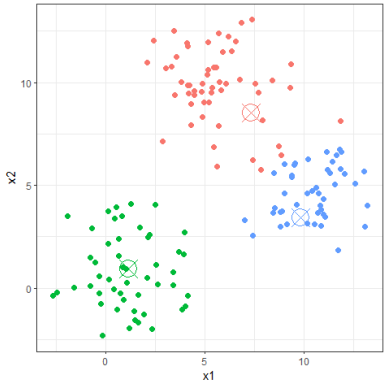

Figure: Iteration 08

Clustering
An example of the K-Means algorithm

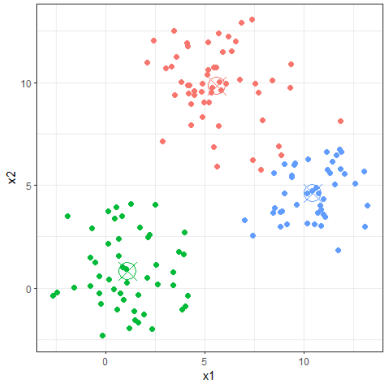

Figure: Iteration 09

Clustering
An example of the K-Means algorithm

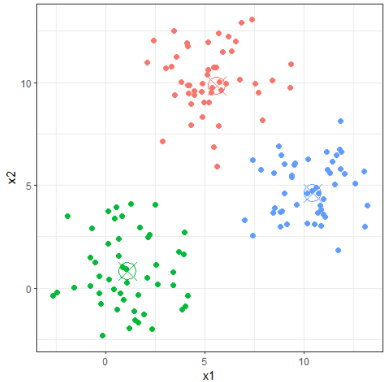

Figure: Iteration 10

Clustering
An example of the K-Means algorithm

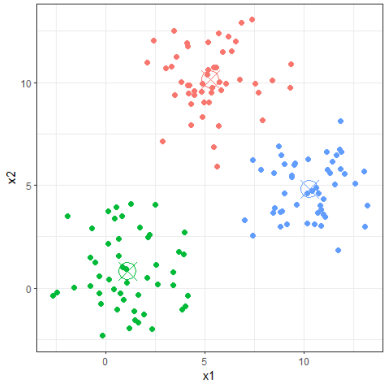

Figure: Iteration 11

Clustering
An example of the K-Means algorithm

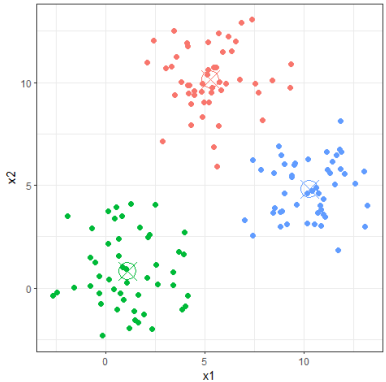

Figure: Iteration 12

Clustering
An example of the K-Means algorithm

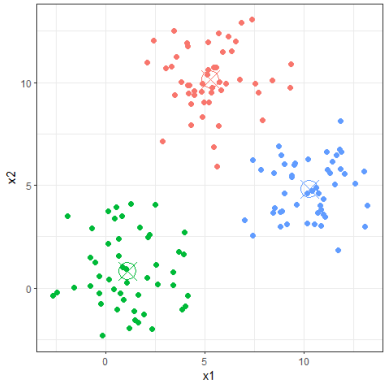

Figure: Iteration 13

Clustering
Hierarchical Clustering

Shrinkage
Lasso

Shrinkage
Ridge Regression

Shrinkage
Elastic Net

Classification
Linear Discriminant Analysis (LDA)

Classification
Quadratic Discriminant Analysis (QDA)

Classification
Naive Bayes

Trees

Classification and Regression Trees (CART)

Trees
Bagging

Trees
Random Forest

Trees
Boosting

Support Vector Machines

Neural Networks